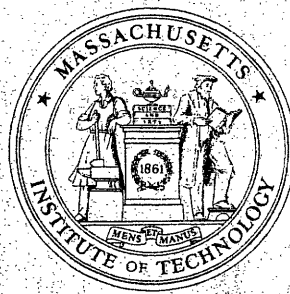


OPERATIONS RESEARCH CENTER

working paper



**MASSACHUSETTS INSTITUTE
OF TECHNOLOGY**

A Nested Decomposition Method
For
Vehicle Routing and Scheduling

by

Anthony O. Lee and Jeremy F. Shapiro

OR 168-87

September 1987

A NESTED DECOMPOSITION METHOD
FOR
VEHICLE ROUTING AND SCHEDULING*

Anthony O. Lee
Jeremy F. Shapiro

Operations Research Center
Massachusetts Institute of Technology
September, 1987

* Research supported in part by a grant from the Charles Stark Draper Laboratory, Inc., Cambridge, Massachusetts.

Abstract

A vehicle routing and scheduling model is constructed in which time dependent Traveling Salesman problems are combined with integer programming sub-models for assigning customers to vehicles. A nested decomposition method based on Benders' decomposition and Lagrangean relaxation for exploiting these structures is presented. Although the method is deficient (non-convergent) due to duality gaps in the Lagrangean relaxations, computational experience, at least for some problems, is promising. Extensions of the basic model to handle time windows and multiple depots are also presented.

A NESTED DECOMPOSITION METHOD FOR VEHICLE ROUTING AND SCHEDULING PROBLEMS

1. INTRODUCTION

Fisher and Jaikumar (1978) presented a mathematical programming model for vehicle routing and scheduling that lends itself to optimization by Benders' (resource directed) decomposition; see also Magnanti (1981). With this decomposition, customers are assigned to vehicles in a Master Model, and each vehicle is routed through the customers assigned to it by solving a Traveling Salesman Problem (TSP) sub-model. The model and optimization approach is appealing, in large part because all feasible vehicle routes are implicitly taken into account. By contrast, the vast majority of packages for practical solution of vehicle routing and scheduling assume that a representative, but non-exhaustive, set of trial routes is somehow generated and passed to a model to be evaluated.

A serious complication with the formulation of Fisher and Jaikumar, however, and with their decomposition approach, is that convergence to an optimal solution for the overall problem is contingent upon achieving linear programming (LP) polyhedral representations of the TSP sub-models. Optimal dual variables for these LP's are used to write Benders' cuts for the Master Model. In principle, Gomory cuts could be added iteratively to inexact LP approximations of the TSP sub-models until TPS tours are obtained. The Benders' cuts would then be derived from these

LP's. This is a daunting prospect and undoubtedly the reason why the models and the decomposition approach for optimizing them was never implemented. Instead, Fisher and Jaikumar (1981) developed a heuristic for approximating the TSP cost structures which they imbedded in a generalized assignment model.

In this paper, we report on a related but different model for the vehicle routing and scheduling problem, and a Benders' decomposition approach for optimizing it. The differences are twofold. First, in the model, we formulate the TSP sub-models as shortest route problems with side constraints (see Picard and Queyranne (1978) and Houck et al (1980)). Second, in the decomposition, we apply Lagrangean relaxation to the TSP sub-models in order to exploit simple shortest route structures. Optimal, or near optimal, Lagrange multipliers are used to write new Benders' cuts for the Master Model. Nested decomposition refers to this concatenation of resource directed and price directed methods. Earlier efforts to develop this approach by Balakrishnan (1982) and Lee and Marge (1985), which led to the approach here, are acknowledged.

Our approach is novel because we allow duality gaps between the TSP sub-models and their duals to occur. For this reason, convergence of the decomposition to an optimal solution to the overall problem cannot be guaranteed. Our approach benefits greatly, however, by being able to employ efficient algorithms for analyzing the TSP sub-models, thereby generating quickly Lagrangean lower bounds, and feasible tours by primal heuristics.

Convergence can only be guaranteed by imbedding the decomposition in a branch and bound scheme.

In the final analysis, a major justification for our method is in its computational performance. Preliminary experience, which we report upon below, indicates that the method works well, at least in some cases. The experience is based on an ac hoc microcomputer implementation that leaves considerable room for improvement. Future experimentation and implementations, hopefully on actual vehicle and routing and scheduling problems, should produce even better results.

We wish also to comment briefly on the poor reputation that Benders' decomposition method has among many researchers. We believe the method's supposed unsatisfactory performance is due largely to an overdependence on mathematical procedures, and a corresponding avoidance of legitimate and effective specific problem solving procedures. In particular, we believe that Benders' decomposition can and should be initiated at an advanced stage, one that follows a systematic analysis of the sub-models. The aim of this analysis is to produce a reasonable initial set of cuts for the Master sub-model.

Our experience with the vehicle routing and scheduling model of this paper, and the stochastic programming model in Bienstock and Shapiro (1987), indicates that, when this is done, a tight bound on the objective function and a satisfactory primal solution are quickly identified. Moreover, preliminary analysis of sub-models is not "cheating", but rather, a sensible approach

to repetitive optimization of a large scale model. Future implementations of Benders' decomposition should recognize this approach, and formalize it by the judicious use of knowledge based systems for initializing and directing the decomposition.

In the following section, we present a statement of the Vehicle Routing and Scheduling Model (VRSM) upon which the nested decomposition method is based. The method is developed in the section after that. We then discuss in detail the Lagrangean shortest route problems, one for each vehicle, that lie at the heart of the computation. A procedure for recovering a complete set of dual variables (Lagrangean multipliers) from each shortest route problem, which are used in writing Benders' cuts, is presented. Implementation and experimental results are discussed in Section 5. Model extensions to handle time windows and multiple depots, are presented in Section 6. The paper concludes with a brief discussion of future research. Readers are referred to Shapiro (1979) for a review of relevant concepts of decomposition and Lagrangean relaxation methods.

2. STATEMENT OF THE MODEL

We present an integer programming model for the basic single depot, multi-vehicle routing and scheduling problem. The goal of this model is to minimize the total cost of operating the fleet of vehicles, while meeting customer demand and vehicle restrictions. The model is as follows:

Indices:

$i = 0, \dots, N$ index of customers (where customer 0 is the starting depot)
 $j = 1, \dots, N+1$ index of customers (where customer $N + 1$ is the ending depot)
 $t = 0, \dots, N$ index of position
 $k = 1, \dots, K$ index of vehicles

Parameters:

c_{ijk} \equiv cost of going from customer i to customer j for vehicle k
 a_i \equiv demand of customer i
 b_k \equiv capacity of vehicle k
 f_k \equiv fixed cost of putting vehicle k on the road

Variables:

y_{0k} \equiv 1 if vehicle is used; 0 otherwise
 y_{ik} \equiv 1 if customer i ($i \geq 1$) is assigned to vehicle k ;
 0 otherwise
 x_{ijkt} \equiv 1 if vehicle k serves customer i in position t
 followed by customer j in position $t + 1$; 0 otherwise

Comments:

1. The starting depot and the ending depot physically correspond to the single depot in our basic problem. They have been differentiated for technical reasons related to optimizing shortest route submodels in our decomposition.
2. Position refers to the order in which a customer is visited.
3. We assume that the cost structure is symmetric and does not depend on the position t .

4. Flexibility in the choice of the x_{ijkt} permitted in the model is possible and desirable. For example, any illogical or uneconomical link between customers i and j can be excluded from consideration by omitting the corresponding variable. For technical reasons, we include the variables x_{iikt} for all i , with cost $c_{iikt} = 0$.

Vehicle Routing and Scheduling Model (VRSM)

$$\begin{aligned}
 v = \min \quad & \sum_{k=1}^K \left[\sum_{j=1}^N c_{0jk} x_{0jk0} + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{N-1} c_{ijk} x_{ijkt} \right. \\
 & \left. + \sum_{i=1}^N c_{i,N+1,k} x_{i,N+1,k,N} \right] \\
 & + \sum_{k=1}^K f_k y_{0k}
 \end{aligned} \tag{1a}$$

Subject to:

$$\sum_{k=1}^K y_{ik} = 1 \quad \text{for } i = 1, \dots, N \tag{1b}$$

$$\sum_{i=1}^N a_i y_{ik} - b_k y_{0k} \leq 0 \quad \text{for } k = 1, \dots, K \tag{1c}$$

For $k = 1, \dots, K$

$$- \sum_{j=1}^N x_{0jk0} + y_{0k} = 0 \tag{1d}$$

$$\sum_{j=1}^N x_{ijj1} - x_{0ik0} = 0 \quad \text{for } i = 1, \dots, N \tag{1e}$$

$$\sum_{j=1}^N x_{ijkt} - \sum_{p=1}^N x_{pik,t-1} = 0 \quad \text{for } \begin{cases} i = 1, \dots, N \\ t = 2, \dots, N \end{cases} \quad (1f)$$

$$\sum_{i=1}^N x_{i,N+1,k,N} - y_{0k} = 0 \quad \text{for } i = 1, \dots, N \quad (1g)$$

$$x_{0jk0} + \sum_{\substack{i=1 \\ i \neq j}}^N \sum_{t=1}^{N-1} x_{ijkt} - y_{jk} = 0 \quad \text{for } j = 1, \dots, N \quad (1h)$$

$$x_{ijkt} \in \{0,1\}; y_{ik} \in \{0,1\} \quad (1i)$$

The objective function (1a) is comprised of the operating costs associated with visiting the N customers and the fixed costs of putting the vehicles on the road. The constraints (1b) state that every customer must be visited exactly once. The constraints (1c) ensure that the capacity b_k of each vehicle is not exceeded.

For $y_{0k} = 1$ the constraints (1d), (1e), (1f) and (1g) describe a shortest route problem connecting the depot (node 0) back to itself (node $N + 1$). The network representation of these constraints is depicted in Figure 1. If, in addition, we consider the assignment constraints (1h) requiring certain customers j to be visited exactly once ($y_{jk} = 1$), we have specified a traveling salesman problem (TSP). In particular, if we define

$$I_k = \{ i \mid y_{ik} = 1, i = 1, \dots, N \} \quad (2)$$

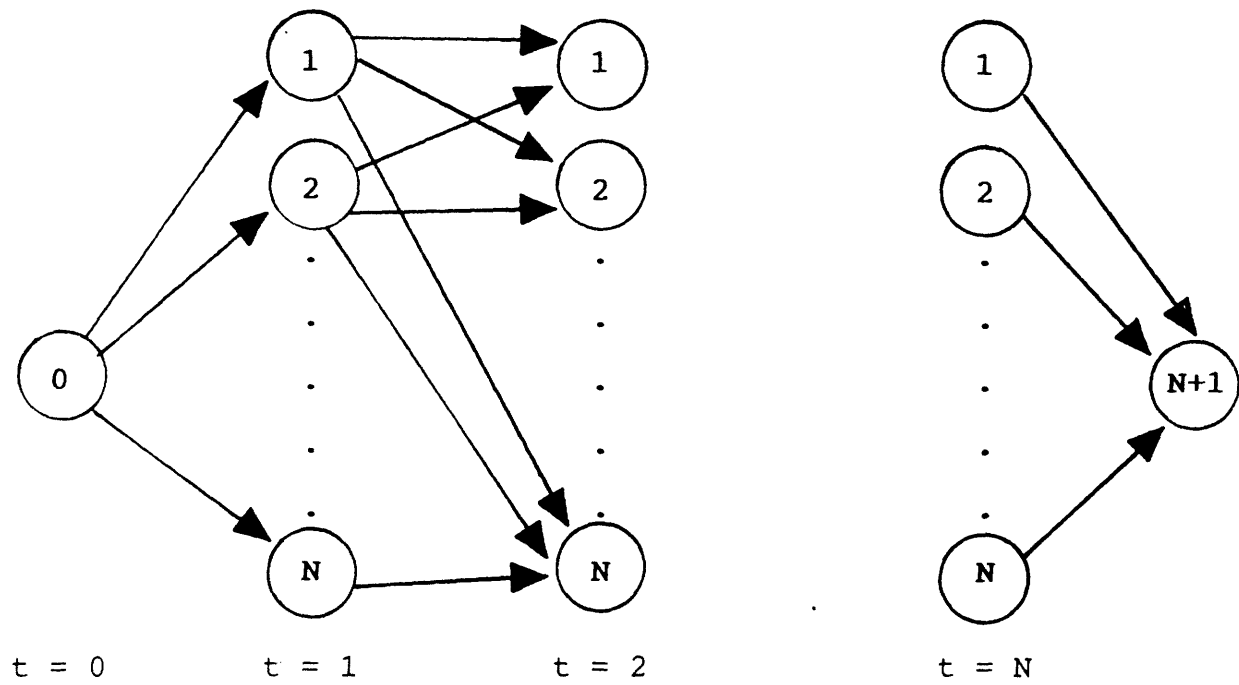
Shortest Route Network for Vehicle k

Figure 1

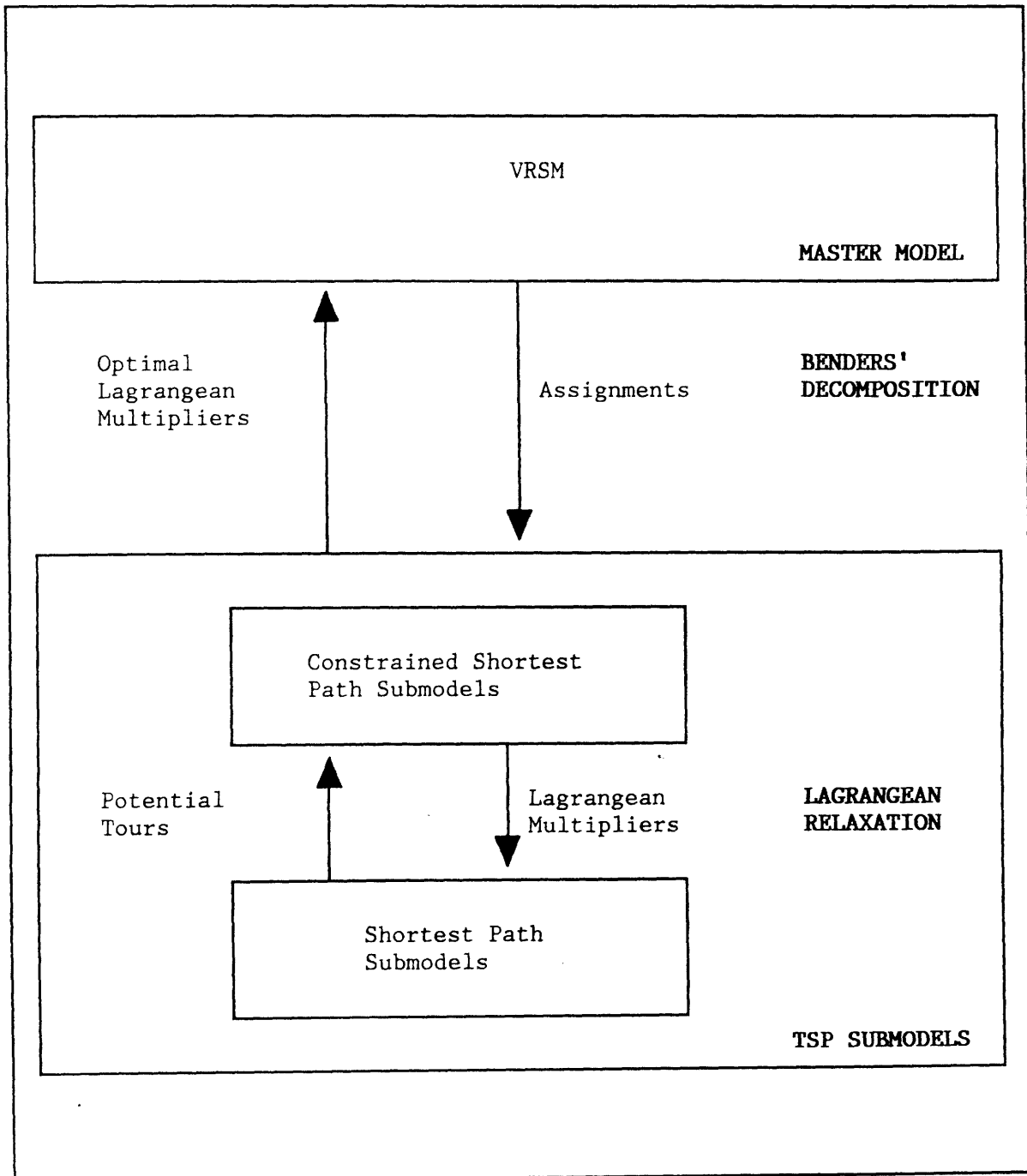
our TSP is to find a minimal length tour starting at the depot node 0 visiting each customer $i \in I_k$ exactly once, and returning to the depot node $N+1$. Since we usually will have $Q_k \equiv |I_k|$ less than N , we use the arcs corresponding to x_{iik} with zero arc length to get from the last customer i visited back to the depot $N + 1$.

Our algebraic statement (1d), (1e), (1f), (1g), and (1h) of the TSP for vehicle k is derived directly from the time dependent traveling salesman problem statement of Fox, Garish and Graves (1980). We do not repeat here their proof that these constraints for each vehicle k correspond exactly to the TSP tours through the customers $i \in I_k$.

3. NESTED DECOMPOSITION METHOD

In this section, we present our nested method as applied to the (VRSM) developed in the previous section. The method is depicted schematically in Figure 2. First, customers are assigned to vehicles by solving the Master Model (MM) comprising constraints (1b), (1c), (1j) and Benders' cuts that approximate the objective function (1a). The optimal objective function value of (MM) provides a lower bound on v , the optimal objective value of (VRSM). The method proceeds by sending the fixed assignments y^* to the submodels.

The resulting submodels are TSP's, one for each vehicle, formulated as constrained shortest path models. To solve the TSP submodels, we apply Lagrangean relaxation to the complicating



Nested Decomposition Method

Figure 2

constraints (1h) to obtain pure shortest path submodels. Subgradient optimization is used to obtain optimal Lagrange multipliers on the complicating constraints. In addition, if a feasible tour is not found from the subgradient optimization, a TSP heuristic determines a feasible tour. Finally, recursions are employed to retrieve missing dual variable that are needed for writing new Benders' cuts. The optimal objective value of the TSP submodels is an upper bound on v , the optimal value objective value of (VRSM).

After the new Benders' cuts are added to (MM), it is resolved. We proceed iteratively until the objective value of (MM) and the TSP submodels approximately converge. To obtain the tightest bound on the difference between the upper and lower bounds, we must always store the best solutions to the TSP submodels that have been found. While the optimal objective values of (MM) are monotone non-decreasing, we have no such guarantee on the objective function values of the TSP submodels. Our development of an algebraic statement of the decomposition approach is based on a matrix form of (VRSM). To this end, define the vectors

$$y^k = (y_{1k}, \dots, y_{Nk})$$

$$\hat{y}^k = (y_{0k}, y_{1k}, \dots, y_{Nk})$$

$$x^k = (x_{ijkt})$$

The re-statement is

$$v = \min \sum_{k=1}^K \{c^k x^k + f_k y_{0k}\} \quad (3a)$$

Subject to:

$$\hat{y}^1, \dots, \hat{y}^K \in Y \quad (3b)$$

For $k = 1, \dots, K$

$$Ax^k - ey_{0k} = 0 \quad (3c)$$

$$Dx^k - y^k = 0 \quad (3d)$$

$$x^k \in \{0,1\}, \hat{y}^k \in \{0,1\} \quad (3e)$$

The constraints (3b) correspond to the constraints (1b) and (1c) in (VRSM), the constraints (3c) to (1d), (1e), (1f), (1g), and the constraints (3d) to (1h). The matrix A is the node-arc incidence matrix of the network depicted in Figure 1 and the $(N+2)$ vector e has two non-zero entries, a $+1$ on the row corresponding to node 0 and a -1 on the row corresponding to node $N+1$. The matrix D extracts the degrees of the vector x^k incidence to the nodes $1, \dots, N$. For simplicity, we have assumed that A does not depend on k .

Consider y fixed at zero-one values satisfying (3b). The residual model is

$$v(\hat{y}^1, \dots, \hat{y}^K) = \sum_{k=1}^K \left\{ f_k y_{0k} + t^k(\hat{y}^k) \right\}$$

where

$$\begin{aligned}
 t^k(\hat{y}^k) &= \min c^k x^k \\
 \text{s.t.} \quad &Ax^k = e y_{0k} \\
 &Dx^k = y^k \\
 &x^k \in \{0,1\}
 \end{aligned} \tag{4}$$

In these terms, the original (VRSM) becomes

$$\begin{aligned}
 v &= \min \sum_{k=1}^K f_k y_{0k} + t^k(\hat{y}^k) \\
 \text{s.t.} \quad &(\hat{y}^1, \dots, \hat{y}^K) \in Y
 \end{aligned} \tag{5}$$

In words, (VRSM) separates into the K TSP's each given by (4).

The nested decomposition method proceeds by dualizing on the degree constraints $Dx^k = y^k$. In particular, for $\mu^k \in \mathbb{R}^N$, define the Lagrangean

$$\begin{aligned}
 L^k(\mu^k \mid \hat{y}^k) &= -\mu^k y^k + \min (c^k + \mu^k D) x^k \\
 \text{s.t.} \quad &Ax^k = e y_{0k} \\
 &x^k \in \{0,1\}
 \end{aligned} \tag{6}$$

Problem (6) is a shortest route model connecting node 0 to node $N+1$ where the arc length for arc (i,j) is given by $c_{ij} + \mu_{jk}$. Of course, $L^k(\mu^k \mid \hat{y}^k) = 0$ if $y_{0k} = 0$.

Letting

$$s^k(\mu^k) = \text{length of a shortest route in (6),}$$

we have

$$L^k(\mu^k \mid \hat{y}^k) = S^k(\mu^k) - \mu^k \hat{y}^k$$

By weak duality we have for all k

$$L^k(\mu^k \mid \hat{y}^k) \leq t^k(\hat{y}^k) \quad (7)$$

and an implied TSP dual problem

$$\begin{aligned} d^k(\hat{y}^k) &= \max L^k(\mu^k \mid \hat{y}^k) \\ \text{s.t. } \mu^k &\in R^N \end{aligned} \quad (8)$$

We use the inequalities (7) as the basis for constructing a Benders' Master Model. In particular, suppose we have previously generated or otherwise selected assignments

$$(\hat{y}^{1,r}, \dots, \hat{y}^{K,r}) \in Y \quad \text{for } r=1, \dots, R$$

Assume we associate a unique $\mu^{k,r}$ with each $\hat{y}^{k,r}$; for example, $\mu^{k,r}$ that is optimal, or close to being optimal, in (8). With this background, we can write the Master Model (MM)

$$\begin{aligned} v^R &= \min \sum_{k=1}^K (f_k y_{0k} + v_k) \\ \text{s.t. } v^k &\geq S^k(\mu^{k,r}) - \mu^{k,r} \hat{y}^k \\ &\quad \text{for } r = 1, \dots, R; \\ &\quad k = 1, \dots, K \\ (\hat{y}^1, \dots, \hat{y}^K) &\in Y \text{ and } \hat{y}^k \in \{0,1\} \text{ for all } k \end{aligned} \quad (9)$$

Theorem 1: The Master Model (g) provides a lower bound on the optimal objective function cost of VRSM (3); that is,

$$v^R \leq v$$

Proof: Let (x^k, \hat{y}^k) for $k = 1, \dots, K$ denote any feasible solution to (VRSM). By the definition of the Lagrangean, for any k and any r ,

$$\begin{aligned} L^k(\mu^{k,r} \mid \hat{y}^k) + \mu^{k,r} y^k &= S^k(\mu^{k,r}) \\ &\leq (c^k + \mu^{k,r} D^k) x^k \\ &= c^k x^k + \mu^{k,r} y^k \end{aligned}$$

where the inequality follows because x^k is feasible in (6), and the last equality follows because $D^k x^k = y^k$. Rearranging terms, we have for all r

$$c^k x^k \geq S^k(\mu^{k,r}) - \mu^{k,r} y^k \quad (10)$$

For each k , let r_k denote the index in problem (9) such that

$$v^k = S^k(\mu^{k,r_k}) - \mu^{k,r_k} y^k$$

Since r is arbitrary, we have from (10)

$$c^k x^k \geq v^k,$$

Since (x^k, y^k) for $k = 1, \dots, K$ is arbitrary, we could assume it is optimal in (VRSM), or

$$v = \sum_{k=1}^K f_k y_{0k} + c^k x^k \geq v^R = \sum_{k=1}^K f_k y_{0k} + v^k. \quad \blacksquare$$

4. LAGRANGEAN RELAXATION OF THE TSP SUB-MODELS

In the previous section, we developed the nested decomposition method for (VRSM). At the heart of the method are shortest route calculations, one for each vehicle, derived from TSP sub-models by Lagrangean relaxation. Our purpose in this section is to discuss this construction in detail.

For convenience, we re-write the generic TSP sub-model, suppressing the index k that distinguishes vehicles. This problem is

$$\begin{aligned} \min \quad & \sum_{j=1}^N c_{0j} x_{0j0} + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{N-1} c_{ij} x_{ijt} \\ & + \sum_{i=1}^N c_{i,N+1} x_{i,N+1,N} \end{aligned} \quad (11a)$$

Subject to:

$$\sum_{j=1}^N x_{0j0} = y_0 \quad (11b)$$

$$\sum_{j=1}^N x_{ij1} - x_{0i0} = 1 \quad i = 1, \dots, N \quad (11c)$$

$$\sum_{j=1}^N x_{ij t} - \sum_{j=1}^N x_{ji, t-1} = 0 \quad \begin{matrix} i = 1, \dots, N \\ t = 2, \dots, N-1 \end{matrix} \quad (11d)$$

$$\sum_{j=1}^N x_{i, N+1, N} = Y_0 \quad (11e)$$

$$x_{0j0} + \sum_{\substack{i=1 \\ i \neq j}}^N \sum_{t=1}^{N-1} x_{ijt} = Y_j \quad j = 1, \dots, N \quad (11f)$$

$$x_{ijt} \in \{0,1\} \text{ for all permissible } i, j, t \quad (11g)$$

In this formulation, the y_j assignment variables are fixed at values of zero or one. Without loss of generality, we assume $y_j = 1$ for $j = 0, 1, \dots, Q$, and $y_j = 0$ for $j = Q+1, \dots, N$.

As we indicated in the previous section, we dualize on the degree constraints (11f), thereby reducing (11) to a shortest route problem. Letting μ_j for $j = 1, \dots, N$ denote the dual variables, we obtain

$$L(\mu \mid Y) = -\sum_{j=1}^N \mu_j Y_j + \min \sum_{j=1}^N (c_{0j} + \mu_j) x_{0j0}$$

$$+ \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{N-1} (c_{ij} + \mu_j) x_{ijt}$$

$$+ \sum_{i=1}^N c_{i,N+1} x_{i,N+1,N}$$

Subject to:

$$x_{ijt} \text{ satisfies (11b), (11c), (11d), (11e), (11g)} \quad (12)$$

The μ_j are to be selected, in principle, by maximizing $L(\mu \mid y)$ for $\mu \in \mathbb{R}^N$.

However, although (12) is defined over all N customers, we can imagine that the μ_j for the customers not assigned to the vehicle ($j = Q+1, \dots, N$) are chosen to be so high that those nodes are never visited. Specific values for these μ_j will be calculated after the others have been selected. The situation is depicted in Figure 3. The Lagrangean optimization (12) is the problem of finding a shortest route from node 0 to node $N+1$ through the upper part of the network where the arc lengths on arcs $(i,t; j,t+1)$ are $c_{ij} + \mu_j$ for $i = 1, \dots, Q$, $j = 1, \dots, Q$ and $t = 0, 1, \dots, N-1$.

The μ_j for $j = 1, \dots, Q$, are iteratively adjusted using subgradient optimization. In particular, letting $x_{ijt}(\mu)$ denote optimal value to the shortest route Lagrangean problem (12), we calculate

$$\sigma_j(\mu) = x_{0j0}(\mu) + \sum_{t=1}^N \sum_{\substack{i=1 \\ i \neq j}}^N x_{ijt}(\mu) - y_j \quad \text{for } j = 1, \dots, Q$$

and update the μ_j by

$$\mu_j^{\text{new}} = \mu_j + \Theta \sigma_j(\mu) \quad \text{for } j = 1, \dots, Q$$

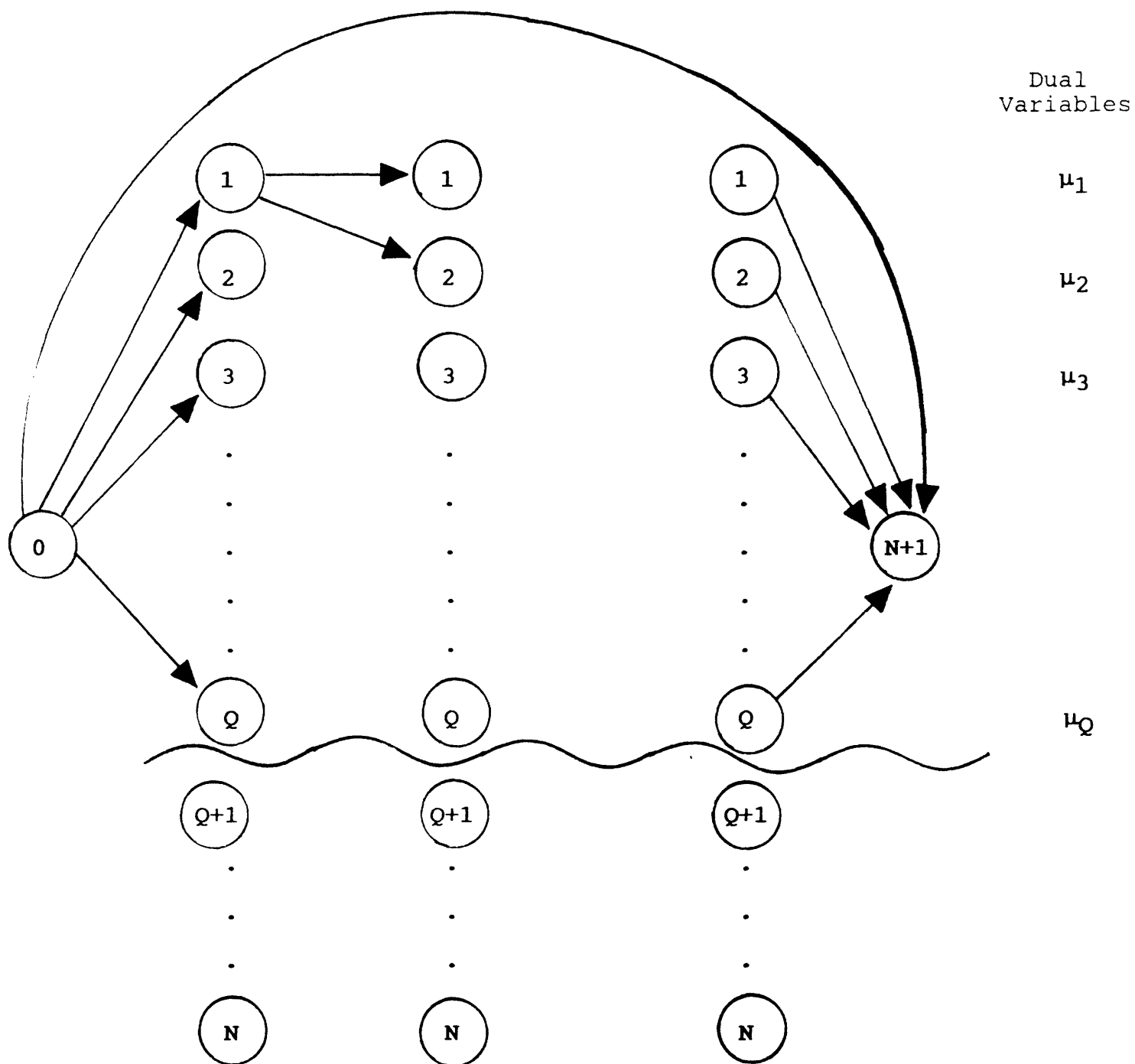
where Θ is determined according to the usual formula of subgradient optimization (see Shapiro (1979)).

Consider now the procedure for retrieving values for the μ_j , $j = Q+1, \dots, N$, after final values for μ_j , $j = 1, \dots, Q$, have been determined. We assume at this point that we have values π_{it} for $i = 1, \dots, Q$, $t = 1, \dots, N$, and π_{00} , such that

$$c_{i,N+1} - \pi_{i,N} \geq 0 \quad \text{for } i = 1, \dots, Q$$

$$\begin{aligned} c_{ij} + \mu_j - \pi_{it} + \pi_{j,t+1} &\geq 0 && \text{for } i = 1, \dots, Q; \\ &&& j = 1, \dots, Q; \\ &&& t = 1, \dots, N \end{aligned} \quad (13)$$

$$c_{0j} + \mu_j - \pi_{00} + \pi_{j,1} \geq 0 \quad \text{for } j = 1, \dots, Q$$



Reduced Shortest Route Network

Figure 3

These are the optimality conditions for the shortest route problem. The π_{it} actually measure the length of the shortest route path from node i,t to node $N+1$. They can be computed recursively starting at node $N+1$ and working backward to node 0.

Recovery of the missing μ_j values proceeds by calculating the π_{it} for $i = Q+1, \dots, N$ and $t = 1, \dots, N$. We still assume that the μ_j for $j = Q+1, \dots, N$ will be chosen sufficiently large that the shortest route from node i,t for $i \geq Q+1$ begins with a transition to a node $j,t+1$ for $j \leq Q$. Specifically, the π_{it} are computed recursively by

$$\pi_{i,N} = c_{i,N+1} \quad \text{for } i = Q+1, \dots, N$$

and for $t = N-1, N-2, \dots, 1$

$$\pi_{i,t} = \underset{j=1, \dots, Q}{\text{minimum}} \{c_{ij} + \mu_j + \pi_{j,t+1}\} \quad (14)$$

$$\text{for } i = Q+1, \dots, N$$

Once all the π_{it} are known, we compute the μ_j for $j = Q+1, \dots, N$ by

$$\mu_j = \max \{ \pi_{it} - c_{ij} - \pi_{j,t+1} \} \quad (15)$$

$$\text{for } i = 1, \dots, N$$

$$t = 0, \dots, N-1$$

Condition (15) ensures that for any node i,t

$$\pi_{it} \leq c_{ij} + \mu_j + \pi_{j,t+1} \quad \text{for } j = Q+1, \dots, N$$

or, in words, that the shortest route from node i, t for any $i = 1, \dots, Q$ to node $N+1$ need not begin with a transition to any node $j, t+1$ for any $j = Q+1, \dots, N$. Thus, the previously computed length $S(\mu)$ of the shortest route remains unchanged, and a complete set of μ_j have been computed. This in turn ensures that a Benders' cut for the Master Model (9) written for this dual vector μ and the given vehicle is valid.

5. IMPLEMENTATION

To implement the nested decomposition procedure, we split it into two distinct phases. First, we used the mixed integer programming code HYPERLINDO/PC (Schrage (1984)) to solve the Master Model (9). Then, we coded a PASCAL program to analyze the TSP submodels (11), where the imbedded shortest route Lagrangean problems (12) were optimized by an algorithm based on a dynamic programming routine taken from Picard and Queyranne (1978). Subgradient optimization was employed to update the multipliers in the Lagrangeans.

For generating primal feasible solutions to the TSP submodels, we used a modified version of the 2-opt TSP heuristic presented in Syslo, Deo and Kowalik (1983). Since our submodel with costs modified by the Lagrange multipliers is generally not symmetric, 2-opt may cycle. Thus, we limited the 2-opt procedure to a maximum number of exchanges.

We analyzed three test problems with the following characteristics:

Problem #1

2 vehicles with capacity 10
6 customers with a total demand of 18
No vehicle fixed costs
Travel costs varying from 19 to 90

Problem #2

3 vehicles with capacities 150, 175, and 200
20 customers with demand varying from 9 to 40 totalling 409
Fixed costs of 400 for each vehicle
Travel costs varying from 15 to 185

Problem #3

3 vehicles with capacities 150
20 customers with demands varying from 5 to 50 totalling 440
No vehicle fixed costs
Same travel costs as problem #2

For more detail about these problems, for example, the distance matrices, see Lee (1987).

A summary of the results of our runs are listed in Table 1. We did not attempt to compute optimal solutions to the test problems. Thus, we know only that the minimal costs lie

somewhere between the lower bounds B and the upper bounds A in Table 1. Comparing the results for problems 2 and 3, we hypothesize that the nested decomposition method produced a much smaller gap for problem 2 than for problem 3 because problem 2 has larger fixed costs and also because problem 3 has a much smaller excess delivery capacity.

Problem Number	Number of Vehicles	Number of Customers	Major Iterations	A Cost of Best Known Feasible Solution	B Greatest Known Lower Bound	Percent Error $\frac{A-B}{A}$
1	2	6	5	266	228.27	.142
2	3	20	4	2197	1851.46	.157
3	3	20	5	1045	610.52	.416

Results

Table 1

The results suggest that the nested decomposition method can produce reasonable results, but considerable more testing is required. Significant modification of the methods are also likely. The computing environment for the implementations just described was slow and awkward to use, and the branch and bound search was not well suited to the application. Any future

experimentation will require optimization routines that are more modular in order to permit flexible testing and assembly of the routines.

Several algorithmic extensions should be tested in future computational experimentation. One is to strengthen the Benders' cuts by deriving them from shortest routes from the origin (starting depot node) to all nodes in the shortest route Lagrangean problem (12), rather than deriving them as we did from shortest routes from the origin only to the destination (ending depot node). This type of procedure for strengthening Benders' cuts is suggested by Magnanti and Wong (1981).

The performance of subgradient optimization in selecting Lagrange multipliers as described in section 4 was somewhat erratic, but difficult to evaluate independently of the other algorithmic procedures employed in the method. An alternative, which is potentially more stable because it maintains a history of previous computation, is generalized linear programming, otherwise known as column generation. For VRSM, the columns being generated correspond to trial routes for each truck. In fact, the application of this method to VRSM yields an interesting new model formulation. The new formulation, and the column generation approach, will be discussed in a forthcoming paper (Lee and Shapiro (1987)).

6. MODEL EXTENSIONS

More complex and realistic vehicle routing and scheduling problems can be modeled and optimized by extensions of the approach presented in this paper. In this section we discuss two important extensions; the incorporation of time windows and multiple depots.

a. Time Windows

The time dependent TSP's used in the constructions above are well suited for modeling delivery time windows. To do so, the index t in the x_{ijkt} variables takes on the time dimension, and transitions from customer i to customer j in general require more than one time period. In particular, in the network of Figure 1, the potential movement of truck k leaving customer i at time t and proceeding to customer j is captured by an arc from node (i,t) to node $(j,t+d_{ijk})$ where d_{ijk} is the travel time for truck k . The travel time may include unloading time at customer j . Thus, vehicle k is ready to travel to some other customer at time $t+d_{ijk}$.

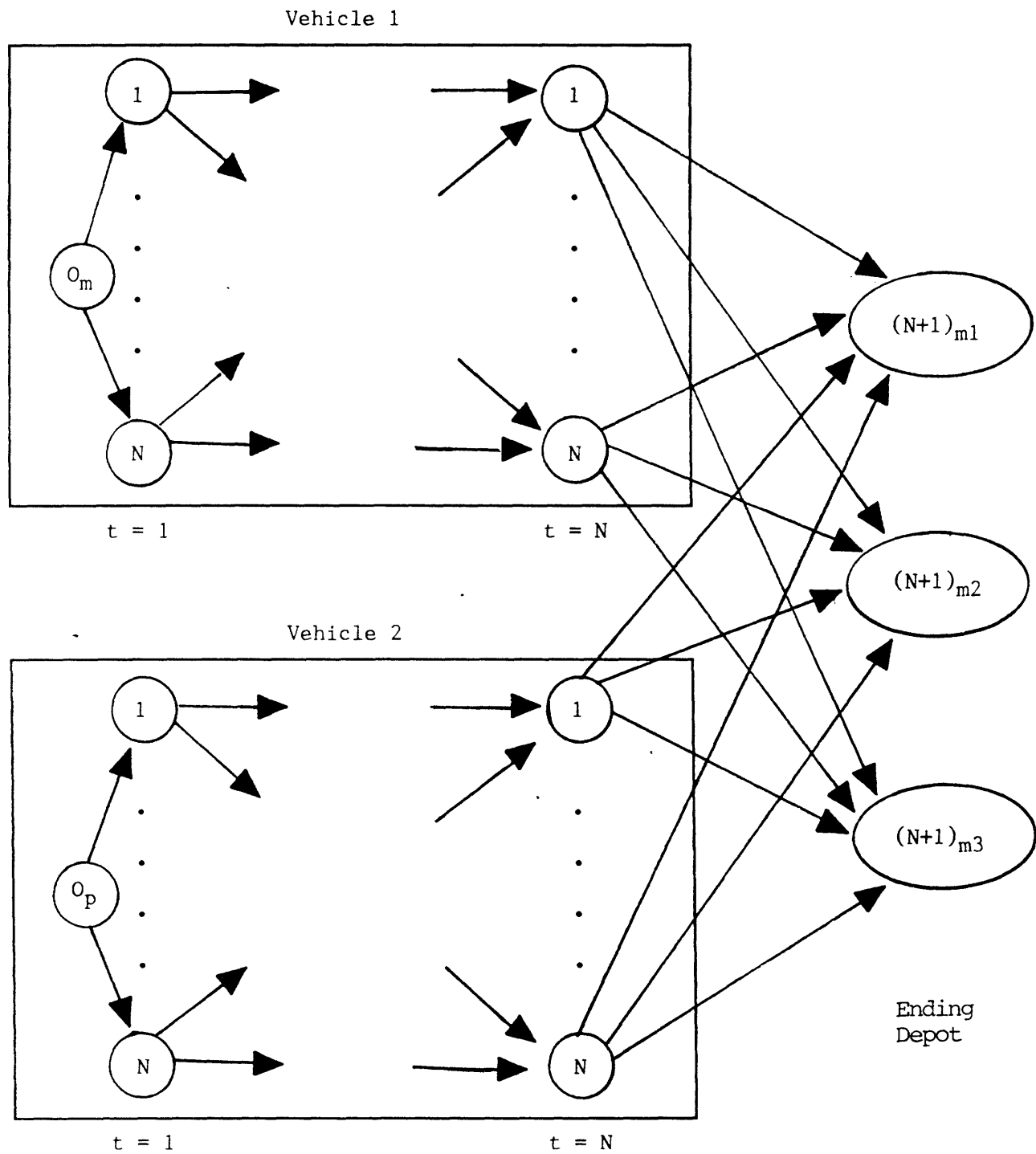
Suppose now that delivery to customer w must take place during $t \in T_w$. This restriction is modeled by including the arcs from (i,t) to $(w,t+d_{iwk})$ only if $t+d_{iwk} \in T_w$. Algebraically, this means that the $x_{iwk}t$ are included in VRSM (model (1)) only for certain t . Moreover, the equations (1e), (1f) and (1g) must be suitably modified to incorporate this restriction and the altered nature of the flow balance equations. Thus, the structure of the TSP's, and the associated shortest route problems is altered.

The structure of the nested decomposition, scheme, however, remains unaltered. We omit further details here.

b. Multiple Depots

This formulation is based on Bodin et al (1983). Suppose there are M depots which we label O_1, O_2, \dots, O_M . Then, in our model we would also have M corresponding ending depots $(N+1)_1, (N+1)_2, \dots, (N+1)_M$. At each origin O_m , we assume there are K_m vehicles available. Moreover, we assume there are upper and lower bounds U_m and L_m imposed on the number of vehicles that end up at depot m (e.g., $U_m = L_m = K_m$).

Figure 4 depicts an extension of the network in Figure 1 which incorporates multiple ending depots. On the left hand side, each truck begins its route at the depot where it is located. On the right hand side, it may go to any one of a number of ending depots. Additional constraints on the flow into each ending depot $(N+1)_m$ are required to ensure that the number of vehicles sent there lies between L_m and U_m . These are bundle constraints that destroy the structure of the nested decomposition scheme. However, the structure can be regained by dualizing on the bundle constraints; in effect, by rewarding or penalizing specific trucks in the TSP's and shortest route calculations on the basis of the depot where they end up.



Modeling Multiple Depots

Figure 4

7. CONCLUSIONS AND AREAS OF FUTURE RESEARCH

In this paper, we have presented a nested decomposition method for vehicle routing and scheduling. As we have shown, the approach involves several novel constructions for analyzing these problems based on the application of Benders' decomposition method to a model with discrete submodels. Computational experience has been encouraging, but more experimentation is needed to evaluate and refine the approach. Several areas of future research are indicated.

First, experimentation in a flexible, more powerful computing environment than that used to date is necessary. The methods discussed in section 5 for writing stronger cuts for the Benders' Master Model should be tested. The column generation formulation for VRSM (1) should also be tested and compared to the methods in this paper. Finally, the decomposition methods should be tested in the context of a branch and bound search for optimal customer assignments to trucks.

An important theoretical result omitted from the development of the nested decomposition method in section 3 is a characterization of duality gaps. In particular, preliminary analysis indicates that the gap between v , the minimal cost of VRSM (1), and the best achievable lower bound from the Master Model (9), can be bounded by a sum of duality gaps between the TSP's (4) and their dual problems (8). Such a theoretical result would have practical implication if we were to discover empirically that the latter duality gaps tend to be small.

Finally, we mention that the modeling and decomposition approach presented here can be applied to other scheduling problems that consist of several TSP's which are synthesized into larger, more complex models. Lee (1987) presents such a formulation of the job shop scheduling problem, and outlines a nested decomposition scheme for optimizing it. The TSP's correspond to sequencing jobs to be performed on a given machine. Another potential application of the approach in this paper is to batch process manufacturing where the TSP's correspond to sequencing and lot sizing of products to be produced on given machines.

REFERENCES

- Balakrishnan, A. (1982), "A Benders' decomposition algorithm for vehicle routing,": Unpublished MIT term paper.
- Bienstock, D. and J. F. Shapiro (1987), "Optimizing resource acquisition decisions by stochastic programming," Management Science, To appear.
- Bodin, L. et. al. (1983), "Routing and Scheduling of vehicles and crews: the state of the art," Computers and Operations Research, 10, 63-211.
- Fisher, M. L. (1981), "The Lagrangean relaxation method for solving integer programming problems," Management Science, 27, 1-18.
- Fisher, M. L. and R. Jaikumar (1978), "A decomposition algorithm for large-scale vehicle routing," Technical Report 78-11-05, Dept. of Decision Sciences, The Wharton School, University of Pennsylvania.
- Fox, K. R. and R. Jaikumar (1981), "A generalized assignment heuristic for vehicle routing," Networks, 11, 109-124.
- Lee, A. O. (1987), "Nested decomposition methods for vehicle routing and scheduling," Unpublished MIT thesis paper.
- Lee, A. O. and J. F. Shapiro (1987), "A column generation model for vehicle routing and scheduling," (in progress).
- Magnanti, T. L. (1981), "Combinatorial optimization and vehicle fleet planning: perspectives and prospects," Networks, 11, 179-213.
- Magnanti, T. L. and R. T. Wong, "Accelerating Benders' decomposition: algorithmic enhancement and model selection criteria," Operations Research, 29, 464-484.
- Marge, C. R. and A. O. Lee (1985), "Solving the vehicle routing and scheduling problem: a Benders' decomposition procedure," Unpublished MIT term paper.
- Picard, J. C. and M. Queyranne (1978), "The time-dependent traveling salesman problem and application to the tardiness problem in one-machine scheduling," Operations Research, 26, 86-110.

Schrage, L. (1984), Linear, integer, and quadratic programming with LINDO, The Scientific Press, Palo Alto, CA.

Shapiro, J. F. (1984), Mathematical programming: structures and algorithms, John Wiley and Sons, New York.

Syslo, M. M., N. Dea and J. S. Kowalik, Discrete optimization algorithms with PASCAL programs, Prentice-Hall, Englewood, NJ.

